THE COMPARATIVE ACCURACY OF DIFFERENT FORMS OF QUINQUENNIAL AGE GROUPS.

BY ALLYN A. YOUNG.

The United States census reports the ages of the persons enumerated by single years. Over one hundred age classes are thus differentiated. It is, of course, impracticable to combine so detailed a classification with the results of more than a few other census inquiries. In the majority of cases in which the census differentiates economic or social classes by age, recourse must be had to age groups. As the importance of age statistics is due in large part to the added value which they give to other results of the census, the form of groups used deserves careful consideration.

From the standpoint of accuracy a classification by groups is preferable to one by single years. Because of the excessive concentration of reported ages on the years which are multiples of 5, a quinquennial grouping is especially accurate.

The numerical accuracy of an age group depends upon the ratio between the number of reported ages which it includes and the number of persons enumerated whose true ages would be included within the same group. If the field of attraction of each age at which concentration occurs does not extend beyond the limits of one age group, the degree of concentration will not affect the accuracy of the age groups. The most accurate form of age group, then, is one which will include with each year of concentration the years from which this concentration is chiefly drawn. The choice of an age group should be influenced by the manner in which concentration takes place. Is it then the prevailing tendency simply to return one's age at the nearest year of concentration? Is there a special tendency to return ages at less than the truth? If there is such a tendency, is it strong enough to affect the comparative accuracy of different

forms of age groups? It is evident that we must be able to answer these questions definitely before we can have a basis for judging the accuracy of a method of age grouping.

If quinquennial groups are used, a method that naturally suggests itself is to place the five lowest ages of the table in the first group, the next five ages in the second group, and so to the end of the table. This method is used by the United States census. There is no reason to believe that its adoption was influenced by any special reasons beyond those of convenience. Criticisms that have been made on this form of grouping have been based on the fact that each group begins with a multiple of 5, that is, with a year of concentration. Mr. George K. Holmes has made the following criticism: —*

It seems generally to be agreed by those who have given the distribution of ages as reported by the enumerators any consideration for the purpose of tabulating them, that such a grouping as this should not be repeated. The years which are notoriously erroneous, that is, the quinquennial and decennial years, are placed at one end in each classification.

.   .   .   .

I would suggest that a quinquennial grouping should be continued, under a new arrangement. the first group to contain ages under three, the next group to be 3 to 7 years. 8 to 12, 13 to 17, and so forth. This would place the especially erroneous years in the middle of the groups in each case.

Mr. Holmes does not state why it is preferable to place the "especially erroneous years in the middle" rather than at the end of the groups. But it is evident that his criticism rests upon the assumption that the prevailing tendency is to return ages at the nearest year of concentration; that is, that approximately as many persons are returned as older than the truth as are returned as younger than the truth, or, in other words, that over statement of age and under statement of age are about equally common.†

* *The Federal Census,* Publications of the American Economic Association, New Series, No. 2, pp. 57, 58.

† As a matter of fact, if all wrongly reported ages were returned at the nearest year of concentration, the number returned as greater than the truth would be slightly larger than the number returned as less than the truth.

It is possible to test this hypothesis and thus to arrive at definite conclusions concerniug the accuracy of different forms of quinquennial age groups.   We have as a starting point the fact that concentration is greater on years which are multiples of 10 than on years which are multiples of 5 and not of 10.   The greater the concentration the greater, of course, will be the depletion in the years from which the concentration is drawn.   The depletion resulting from concentration on the even multiples of 5 will be greater than that resulting from concentration on the odd multiples of 5.   Now, as already stated, the most accurate form of quinquennial age group is one which includes with each year of concentration the years from which the concentration is mainly drawn. If, however, the system of grouping is such that each group excludes some of the years of depletion which have contributed mainly to the year of concentration in that group, and includes some of the years of depletion that have contributed to the year of concentration in the next group, the effect will be to enlarge the groups containing the multiples of 10 at the expense of the other groups.   In other words, if the quinquennial groups are so formed that the fields of attraction of the years of concentration extend beyond the limits of the several groups, a year which is a multiple of 10 would draw relatively more from the neighboring groups than would a year which is an odd multiple of 5.   This would result in making the groups alternately too large and too small.   The excess of the number of reported ages in groups containing years that are even multiples of 5 over the number reported in groups containing years which are odd multiples of 5 is a measure of the comparative inaccuracy of a system of age grouping.   This point may be brought out more clearly by an illustration.

If Mr. Holmes' hypothesis that the erroneously reported years are drawn equally from above and below the year of concentration is correct, an age group containing the years 43–47 should be more accurate than the group 45–49,

although both groups contain the year of concentration 45. The years 48 and 49 would have been drawn upon to make up the large concentration of the year 50, and would therefore be depleted to a greater extent than would the years 43 and 44. The group 43–47 would therefore (on Mr. Holmes' hypothesis) be *relatively** larger than the group 45–49. In this case also the group 48–52 would be *relatively* smaller than the group 50–54.† Evidently the excess of the concentration on the multiple of 10 over that on the odd multiple of 5 has less effect on the relative sizes of the groups which we have assumed to be the more accurate than on the other form of groups. The principle thus illustrated may be expressed as follows : —

Of two series of quinquennial age groups, the more accurate will be the one in which the alternate excess and deficiency in the size of the successive groups is less marked.

To apply this principle to the census age returns, we may represent the relative size of each age group by the expression $\dfrac{G_x}{G_{x-1} + G_{x+1}}$.‡ Here G_x represents the number enumer-

---

* By the *relative* inequality of age groups is here meant an inequality distinct from the natural inequality caused by the fact that as age increases population decreases.

† This illustration may be presented more clearly by the use of mathematical symbols. In the following equations the signs of inequality are used as referring to *relative* and not to absolute inequality : —

Let $P_x =$ number reported at any age x.

Let $G_5 =$ number reported in group containing multiple of 5.

Let $G_{10} =$ number reported in group containing multiple of 10.

If $G_5 = P_{43} + P_{44} + (P_{45} + P_{46} + P_{47})$

and $G'_5 =$ $(P_{45} + P_{46} + P_{47}) + P_{48} + P_{49},$

$G_5 - G'_5 = (P_{43} + P_{44}) - (P_{48} + P_{49})$

Since $(P_{43} + P_{44}) > (P_{48} + P_{49}), G_5 > G'_5.$ (A)

Similarly, if

$G_{10} = P_{48} + P_{49} + (P_{50} + P_{51} + P_{52})$

and $G'_{10} =$ $(P_{50} + P_{51} + P_{52}) + P_{53} + P_{54}$

Since $(P_{48} + P_{49}) < (P_{53} + P_{54}), G_{10} < G'_{10}.$ (B)

From (A) and (B) we conclude that

$(G_{10} - G_5) < (G'_{10} - G'_5).$

‡ This expression represents the numerical relation of each age group to the sum of the two neighboring groups. For our purposes such an expression is much preferable to one which represents the numerical size of an age group in its relation to the entire number of ages reported.

ated in any age group x. The following table illustrates the application of this expression and also shows the true relations between various age groups in a normal stationary population. For convenience the results are expressed in percentages : —

TABLE SHOWING PROPORTION OF NUMBER IN EACH AGE GROUP TO THE SUM OF THE NUMBERS IN THE TWO NEIGHBORING GROUPS. BASED ON FARR'S LIFE TABLE NO. 3, l_x COLUMN.

Age Group.	G_x	$G_x-1 + G_x+1$	$\dfrac{G_x}{G_x-1 + G_x+1}$ Expressed as a Per Cent.
0–4	4,166,056	.. ....	....
5–9	3,602,611	7,641,334	47.1
10–14	3,475,278	7,185,525	48.4
15–19	3,382,914	6,732,376	50.2
20–24	3,257,098	6,493,061	50.2
25–29	3,110,147	6,212,843	50.1
30–34	2,955,745	5,904,065	50.1
35–39	2,793,918	5,578,370	50.1
40–44	2,622,625	5,232,548	50.1
45–49	2,438,630	4,860,278	50.2
50–54	2,237,653	4,445,063	50.3
55–59	2,006,433	3,969,355	50.5
60–64	1,731,702	3,413,225	50.7
65–69	1,406,792	2,768,339	50.8
70–74	1,036,637	2,065,137	50.2
75–79	658,345	1,374,418	47.9
80–84	337,781	787,925	42.9
85–89	129,580	371,759	34.8
90–94	33,978	135,081	25.2
95–99	5,501	......	...

It will be noted that during the periods of youth and middle age, when the death rate is fairly constant, the number reported in each age group is slightly in excess of the arithmetic mean of the numbers reported in the two neighboring groups. It is very improbable that an accurate table of the ages of an actual population, constituted as is the population of the United States, would exhibit any such regularity in the numerical relations of the different age groups. But if we discover a *uniform irregularity* in the

returns for different classes of the population we may safely conclude that such an irregularity is due to the manner in which ages are reported and not to the peculiar age composition of the population.

As a preliminary to the application of this test to the census returns, the ages of the population of the United States, as reported in 1890, have been grouped by three methods. The census plan is followed in the first method, a year of concentration being placed at the beginning of each group. The second method is the one suggested by Mr. Holmes, the third or median year of each group being a year of concentration. In the last method the year of concentration is placed at the end of the group. For convenience these methods will be referred to as Method I, Method II, and Method III, respectively. The following tables show the comparative numerical values of the age groups formed by these three methods for different classes of the population: —

PERCENTAGE RELATION BETWEEN THE NUMBER REPORTED IN EACH SPECIFIED AGE GROUP AND THE NUMBER REPORTED IN THE TWO NEIGHBORING AGE GROUPS, WHEN THE FIRST YEAR OF EACH GROUP IS A YEAR OF CONCENTRATION.

[*Method I.*]

Age Groups.	Aggregate Population.			Native White.		Colored.	
	Total.	Male.	Female.	Total.	Female.	Total.	Female.
5–9	51.6	51.3	51.9	58.0	58.3	52.5	53.0
10–14	49.8	50.5	49.0	49.2	48.6	52.5	51.1
15–19	49.6	48.6	50.5	50.9	51.7	49.3	50.6
20–24	52.6	52.2	53.0	51.9	51.9	51.1	51.7
25–29	48.5	48.8	48.2	46.2	46.4	49.2	48.9
30–34	50.3	51.1	49.5	52.1	51.7	43.1	42.1
35–39	49.8	50.3	49.3	49.3	48.6	56.6	55.8
40–44	48.3	47.7	48.9	47.4	47.9	45.1	47.1
45–49	49.6	49.5	49.6	49.0	49.4	50.9	50.0
50–54	52.8	52.8	52.8	51.8	51.8	55.6	55.4
55–59	44.2	44.3	44.1	47.8	44.9	38.0	36.2
60–64	54.3	54.3	54.4	52.2	52.1	59.9	62.2
65–69	46.8	46.8	46.7	47.7	47.9	43.4	41.7
70–74	50.0	50.2	49.8	50.7	50.1	51.2	54.2
75–79	43.4	43.1	43.7	44.8	45.3	40.1	39.0
80–84	43.5	42.0	45.1	44.1	45.0	50.8	55.7
85–89	33.1	31.6	34.4	33.4	35.3	32.3	31.3
90–94	29.2	27.2	30.9	25.9	27.2	47.9	51.8

PERCENTAGE RELATION BETWEEN THE NUMBER REPORTED IN EACH SPECIFIED AGE
GROUP AND THE NUMBER REPORTED IN THE TWO NEIGHBORING AGE GROUPS,
WHEN THE MEDIAN YEAR OF EACH GROUP IS A YEAR OF CONCENTRATION.

[*Method II.*]

Age Groups.	Aggregate Population.			Native White.		Colored.	
	Total.	Male.	Female.	Total.	Female.	Total.	Female.
8–12	49.2	49.6	48.8	48.3	48.0	51.6	50.8
13–17	49.1	49.0	49.2	50.0	50.1	49.9	50.1
18–22	52.4	51.4	53.4	52.9	53.5	50.9	51.9
23–27	48.8	48.8	48.7	46.3	46.3	50.3	50.8
28–32	52.3	53.1	51.3	52.6	52.2	47.1	45.7
33–37	47.0	47.5	46.5	48.3	47.7	45.4	44.7
38–42	52.1	51.9	52.2	49.9	49.8	58.3	60.2
43–47	46.2	45.9	46.4	46.8	47.4	39.8	39.3
48–52	53.5	52.9	54.1	51.0	51.7	64.6	66.4
53–57	46.9	47.6	46.1	48.7	47.8	37.6	34.2
58–62	49.7	49.0	50.4	47.4	47.7	58.3	63.1
63–67	50.4	51.0	49.8	51.5	51.4	42.8	39.6
68–72	48.6	48.4	48.7	48.4	48.1	54.6	58.8
73–77	45.3	45.2	45.4	47.3	47.6	37.8	36.6
78–82	44.5	43.8	45.1	44.6	44.7	54.0	58.6
83–87	37.3	35.2	39.3	38.9	41.4	32.0	31.6
88–92	31.0	29.8	32.1	28.3	29.2	51.2	53.7

PERCENTAGE RELATION BETWEEN THE NUMBER REPORTED IN EACH SPECIFIED AGE
GROUP AND THE NUMBER REPORTED IN THE TWO NEIGHBORING AGE GROUPS,.
WHEN THE LAST YEAR OF EACH GROUP IS A YEAR OF CONCENTRATION.

[*Method III.*]

Age Groups.	Aggregate Population.			Native White.		Colored.	
	Total.	Male.	Female.	Total.	Female.	Total.	Female.
6–10	52.2	52.2	52.1	51.6	51.6	57.8	54.5
11–15	48.4	49.1	47.7	48.5	47.9	49.5	48.1
16–20	50.7	48.8	52.8	51.8	53.5	50.5	53.5
21–25	50.9	51.5	50.3	49.7	48.9	50.7	49.6
26–30	52.5	52.8	52.1	50.4	50.7	52.0	52.0
31–35	45.5	46.3	44.6	47.7	47.0	38.4	36.8
36–40	54.3	54.1	54.6	52.2	52.1	66.2	68.2
41–45	44.2	44.2	44.3	44.8	44.8	36.5	36.7
46–50	55.5	54.6	56.5	53.0	54.2	67.3	69.0
51–55	44.5	45.3	43.6	46.0	45.2	36.4	33.2
56–60	53.4	52.7	54.3	51.2	51.5	62.8	66.6
61–65	47.2	47.7	46.4	48.1	47.6	40.1	38.0
66–70	51.2	50.6	51.8	50.4	50.9	56.5	59.0
71–75	44.8	45.4	44.1	47.1	46.4	37.7	37.1
76–80	47.5	46.6	48.5	47.3	48.0	55.9	59.0
81–85	37.1	36.0	38.2	39.6	40.9	30.8	31.1
86–90	37.9	35.9	39.7	35.2	37.1	56.8	58.5

Considering the results for the total aggregate population, it appears that by each method of grouping, in the period of middle age, there is an excess in every group containing a year which is a multiple of 10, with the exception of the group 40-44. In the following tables the expression obtained for the numerical values of the groups containing multiples of 10 and of the groups containing odd multiples of 5 are placed in parallel columns. The groups used contain the years of concentration from 20 to 55 : —

RELATIVE SIZES OF THE QUINQUENNIAL GROUPS FORMED FROM THE AGE RETURNS OF THE TOTAL AGGREGATE POPULATION BY THREE METHODS OF GROUPING.

Years of Concentration.	Method I.		Method II.		Method III.	
	Groups Containing Multiples of 10.	Groups Containing Multiples of 5.	Groups Containing Multiples of 10.	Groups Containing Multiples of 5.	Groups Containing Multiples of 10.	Groups Containing Multiples of 5.
20	52.6	....	52.4	....	50.7	....
25	....	48.5	....	48.8	....	50.9
30	50.3	....	52.3	....	52.5	....
35	...	49.8	....	47.0	....	45.5
40	48.3	....	52.1	....	54.3	....
45	....	49.6	....	46.2	....	44.2
50	52.8	....	53.5	....	55.5	....
55	....	44.2	....	46.9	....	44.5
Average	51.0	48.0	52.6	47.2	53.2	46.3

The difference between the average numerical values of the two kinds of groups shown in the foregoing table is, by Method I, 3.0; by Method II, 5.4; by Method III, 6.9. The following table shows the corresponding averages and differences for various classes of the population.

It appears from this table that the differences are uniformly smallest by Method I and largest by Method III. Remembering that these differences represent the relative inaccuracy of these different methods, we may conclude definitely that for the ages under consideration, *groups beginning with a*

COMPARISON OF THE DIFFERENCES BETWEEN THE NUMBER OF REPORTED AGES IN
QUINQUENNIAL GROUPS CONTAINING AN EVEN MULTIPLE OF 5 AND IN GROUPS
CONTAINING AN ODD MULTIPLE OF 5, AS SHOWN BY THREE METHODS OF
GROUPING.

Classes of Population.	Multiples of 5.	Method I.	Method II.	Method III.
Total aggregate........	Even	51.0	52.6	53.2
	Odd	48.0	47.2	46.3
	Difference	3.0	5.4	6.9
Aggregate males.......	Even	51.0	52.3	52.6
	Odd	48.2	47.5	46.8
	Difference	2.8	4.8	5.8
Aggregate females....	Even	51.1	52.8	54.0
	Odd	47.8	46.9	45.7
	Difference	3.3	5.9	8.3
Total native white.....	Even	50.8	51.6	51.9
	Odd	48.1	47.5	47.1
	Difference	2.7	4.1	4.8
Native white females ..	Even	50.8	51.8	52.6
	Odd	47.3	47.3	46.5
	Difference	3.5	4.4	6.1
Total colored	Even	48.7	55.2	59.0
	Odd	48.7	43.3	40.5
	Difference	0.0	11.9	18.5
Colored females........	Even	48.6	58.5	60.7
	Odd	47.7	42.3	39.1
	Difference	0.9	16.2	21.6

*year of concentration are more accurate than groups in which
the year of concentration is the median or the last year.*   In

other words, the tendency to return ages as less than the truth is much stronger than the tendency to return them as greater than the truth. The strength of this tendency varies greatly with the different classes of the population. This is shown by a comparison of the excess of the differences by Method III over the differences by Method I for different classes. This excess is least for the total native white (2.1) and greatest for the colored females (20.7). In general, the tables show that the tendency to report ages as less than the truth is stronger with females than with males and with the colored than with the white population.

An examination of the relative sizes of the differences for different classes of the population throws light on the question of the excessive concentration at 10-year intervals. This tendency to report ages at years which are even multiples of 5, rather than at odd multiples, is greater with females than with males and is greater with the aggregate population than with the native white population. The results of two methods of grouping indicate that it is stronger with the colored than with the white population. But by Method I the differences for the total colored and the colored females are only 0.0 and 0.9, respectively,— smaller than the corresponding differences for the native whites or the aggregate population. Reference to the table on page 32 shows that these low differences are due to the excessive size of the age group 35–39, as compared with the two neighboring age groups. The groups containing the year 35, formed by the two other methods of grouping, do not exhibit this peculiarity. On the other hand, there is a marked concentration in each group containing the year 40, except the group formed by Method I (40–44). It would appear that the four years below 40 (36–39) are excessively large as compared with the years above 40 (41–44). This is probably chargeable in part to a peculiar tendency on the part of those whose ages are greater than 40, to return themselves as less than 40. It

would seem also that the concentration on the year 40 is drawn in but very slight degree from years less than 40. This peculiarity, especially noticeable with respect to the colored population, is also traceable in the returns from the native whites and the aggregate population.

In the light of the foregoing explanation of the only part of the table that seems to be contradictory, we may conclude that the tendency to concentrate on the even, rather than on the odd multiples of 5, is much stronger with the colored than with the white population.

It should be noticed that the figures which have been presented do not afford a basis for determining the percentage of inaccuracy of any age group or method of age grouping. The method used is valid only for comparative purposes. Neither is it practicable to apply this method to other than quinquennial groups, as its basis is the excess of concentration on the even multiples of 5 over that on the odd multiples of 5. A decennial group, for example, must contain both an even and an odd multiple of 5, and a series of such groups cannot be studied to advantage by this method. The conclusions which have been reached as to the direction of concentration would seem to have some bearing on the question of the relative accuracy of various forms of decennial age groups.

The method used has another limitation in being applicable to the study of only a portion of the age table. Above 60 years the increase of the death rate is so rapid that a study of the comparative sizes of age groups can hardly lead to trustworthy conclusions. But this part of the population is comparatively small and, from an economic standpoint, comparatively unimportant. It seems probable, however, that the generally accepted opinion, that persons advanced in years very commonly overstate their ages, is entirely correct. In the early years of life also there are rapid changes of the death rate. As the age returns show no marked con-

centration on multiples of 5 below 20, it has been thought best in this study to disregard these lower ages.

SUMMARY OF CONCLUSIONS.

1. More persons return themselves as younger than they are than as older than they are. This is true for both sexes and for all classes of the population. This tendency is strongest in the colored population and is least marked in the native white population. For all classes the tendency is more noticeable in women than in men.

2. On account of the tendency to report ages as less than the truth the years of concentration are preferably placed at the beginning of the quinquennial groups.

3. A number of persons who are over 40 years old report themselves as under 40. This is true for all classes, but is especially noticeable with the colored population. It also seems probable that very few persons less than 40 years old report themselves as 40.

4. The results of this study throw some light on the question of the relative desirability of a quinquennial or some other system of grouping. The English census of 1891 used decennial groups, explaining that " if ages be grouped by given quinquennial periods the figures for the successive quinquennia are too high or too low, accordingly as they include or do not include a year that is an exact multiple of 10. Therefore . . . it is better for such purposes as require accuracy . . . to make use preferentially of decennial periods, and to arrange them in such a way that the year which is a multiple of 10 shall come in the middle of the period ; that is to say, to use such age periods as 25–35, 35–45, etc., rather than 30–40, 40–50, etc." *

It has been shown that the accuracy of a series of quinquennial groups depends very largely on the form of group used. It does not seem that the inaccuracy of the form of

---

* *Census of England and Wales*, 1891, General Report, p. 27.

quinquennial groups used in the United States census is sufficient to warrant the substitution of the less convenient decennial groups.   The choice of the form of decennial group used in the English census was evidently made on the assumption that the wrongly reported ages are drawn equally from above and below the year of concentration.   It has been shown that, for the United States at least, this assumption has no basis in fact.   There is room for considerable doubt as to whether the decennial groups used in the English census are more accurate than the quinquennial groups used in the United States census.